Optimization of Stereo-matching Algorithms Using Existing DEM Data

D.G. Milledge, S.N. Lane, and J. Warburton

Abstract

Here we present a new method for using existing Digital Elevation Model (DEM) data to optimize performance of stereo-matching algorithms for digital topographic determination. We show that existing DEM data, even those of a poor quality (precision, resolution) can be used as a means of training stereo-matching algorithms to generate higher quality DEM data. Existing data are used to identify and to remove gross surface errors. We test the method using true vertical aerial imagery for a UK upland study site. Results demonstrate a dramatic improvement in data quality even where DEM data derived from topographic maps are adopted. Comparison with other methods suggests that using existing DEM data improves error identification and correction significantly. Tests suggest that it is applicable to both archival and commissioned aerial imagery.

Introduction

Digital Elevation Models (DEMs) generated using automated stereo-matching are susceptible to gross error resulting from incorrect identification of homologous point pairs. Labeled as gross errors, they are commonly handled in one of three ways. The first applies some form of filter that smoothes the acquired data. This does not require labeling of individual data points as erroneous but suffers as localized surface errors can propagate into correct elevations (Lane *et al.*, 2000). The second seeks to identify individual erroneous points and to replace them with correct data points, based upon interpolation from surrounding data points that are thought to be correct (e.g., Felicismo, 1994; Westaway *et al.*, 2003; Lane *et al.*, 2004). The main problem with this approach is that errors rarely take the form of readily identifiable spikes that are clearly detected by filters such as those based upon local variance (e.g., the Chauvenet criterion; Taylor, 1997). Commonly, errors are clustered and, particularly when local topographic variance is high, identifying erroneous points is difficult. The third approach revisits the stereo-matching process itself and aims to regenerate the elevation data after optimization of parameters that control the stereo-matching algorithm performance. The main problem with this approach is that additional information is required in order to identify an optimum parameter set. Three options are available to address this. The first uses independently acquired survey data, often viewed as the only independent means of establishing DEM reliability (e.g., Torlegård, 1986; Chandler, 1999; Westaway *et al.*, 2003). However, such data are not always available and this approach undermines one of the key advantages of remote sensing namely that it should involve remote data acquisition as far as is possible. A second option is to identify those points that appear to be most sensitive to changes in parameter values during stereo-matching; Gooch and Chandler (2001) labeled this the "Failure Warning Model." This method has been successfully used to identify data points with a high sensitivity to parameter values and in which there should be less confidence in their reliability (Lim *et al.*, 2005; Yanites *et al.*, 2006). A third option involves generation of a coarser DEM which can then be used to check a finer DEM following the observation that coarser resolution DEMs commonly contain lower percentage point errors (Westaway *et al.*, 2003; Lane *et al.*, 2004).

In this paper, we develop and test an alternative approach to the generation of optimum parameter sets in stereo-matching algorithms. It is based upon the premise that, in many cases, digital elevation model data is already available for an area, commonly held by regional or national agencies (e.g., the global elevation dataset collected by NASA's SRTM mission (Rabus *et al.*, 2003)). These can be used: (a) to identify data points that are likely to be erroneous; (b) to quantify the associated error; and (c) to adjust stereo-matching parameters so as to optimize the associated DEM quality. The existing data are commonly of a degraded precision and resolution as compared with the data generated by automated stereo-matching. We argue that, this aside, such data still retain important information that can be used to generate more precise and better resolution data sets using stereo-matching algorithms. Our test also compares the use of existing data in DEM optimization with other approaches to error identification.

Hypotheses to be Tested

We address four key hypotheses. First, we test the hypothesis that existing DEM data can be used to generate better parameter sets during stereo-matching than those expected for a given terrain type and hence produce higher quality digital elevation data. We label this "Extant Optimization" (optimization using existing elevation data). In testing this hypothesis, we recognize that many stereo-matching algorithms come with both a default parameter set as well as a series of parameter sets optimized for particular terrain types. In order to see if using Extant Optimization is beneficial, we evaluate our results with respect to those that can be obtained using a terrain sensitive parameter set. Second,

Department of Geography, Durham University, Science Laboratories, South Road, Durham, DH1 3LE (d.g.milledge@durham.ac.uk).

Photogrammetric Engineering & Remote Sensing
Vol. 75, No. 3, March 2009, pp. 323–333.

0099-1112/09/7503–0323/$3.00/0

imagery will be of variable radiometric quality dependent on its source and resolution. Hence, we check a second hypothesis that Extant Optimization can be used to compensate for situations where image information content is poor, as might be associated with archival imagery. Third, as there will be variability in the quality of existing DEM data available between different geographical locations we test the third hypothesis that the effectiveness of Extant Optimization depends on the type of DEM data available. Finally, we compare the Extant Optimization method with two other means of identifying DEM error: (a) the Failure Warning Model of Gooch and Chandler (2001), and (b) the coarse DEM method of Westaway *et al.* (2003) and Lane *et al.* (2004).

Methodology

One of our key aims is to optimize the standard area-based image correlation algorithm that has been the subject of study in a number of recent research papers (e.g., Butler *et al.*, 1998; Gooch *et al.*, 1999; Lane *et al.*, 2000; Gooch and Chandler, 2001). We do this for a test site that comprises an area of 1 km^2 of complex upland topography in the English Lake District (longitude: −3.1919°; latitude: 54.6026°; elevation 200 to 700 m above sea level). It is a characteristic steep unforested upland environment with slopes ranging from 0° to 55° and an average slope of 30°. The area has a covering of: Bracken (*Pteridium Aquilinum*), Heather (*Calluna Vulgaris*) and Cotton Grass (*Eriophorum Vaginatum*) with patches of rushes (*Juncus Effusus*) and *Sphagnum* mosses in wetter areas, and some isolated small native deciduous trees. The site bounds a range of topographic features including: gullies, scree slopes, broad and tight topographic hollows, and two different streams of different orders and character. Forested areas were avoided since the acquisition of a "bare earth" model from these areas presents a separate problem, not directly addressed here.

Generation of DEM Data

We undertook a standard approach to DEM generation using stereo-matching on two sets of aerially acquired imagery (Table 1). We chose two sets in order to test the hypothesis that parameter optimization can be a means of compensating for poorer image content, whether due to lower resolution or other effects. Both sets were true vertical color images from metric cameras mounted on aircraft. The 2005 photos were commissioned from the University of Cambridge Unit for Landscape Modeling (ULM) as part of an inventory of landslides from a high-magnitude storm event in the region. Archival imagery for 2004 was obtained from INFOTERRA to provide a pre-failure base line for the same project. Ground control points were collected by surveying features clearly visible on both sets of photographs using a differential GPS (precision of ± 50 mm). The photographs were scanned from diapositives using photogrammetric scanner before being processed in the Leica Photogrammetry Suite (LPS). The extracted DEMs were referenced in the Ordnance Survey (OS) grid with heights above Newlyn datum. The interior orientations of the images were established with an error close to ±7 μm. The exterior orientations were calculated using standard least squares block bundle adjustment with a standard deviation of unit weight of ±1 pixels in both cases. Butler *et al.* (1998) suggest that this is desirable as it neither under nor over constrains the bundle adjustment. All DEMs were extracted at a 1 m resolution.

Existing DEM Data

We identified three sources of DEM data that could guide Extant Optimization (Table 2). These were chosen to represent three data availability scenarios. The first, IFSAR, reflected the situation where a national mapping agency has commissioned and made available high-resolution DEM data using the current generation of synthetic aperture radar systems. Such a situation is likely to become increasingly common, with complete coverage of the United States and Western Europe by the end of 2008 (Intermap, 2007). The second (Ordnance Survey) was chosen to reflect the more common situation where a national mapping agency holds traditional map-based data, normally collected using either analogue or analytical photogrammetry. The third (SRTM) was chosen to reflect globally available data sources that could be used in any DEM generation situation. All three datasets were interpolated using bilinear interpolation to match the 1 m resolution of the generated DEM data.

Assessment of Generated DEMs

Each generated DEM was compared with each existing DEM in order to quantify gross errors. Gross errors were defined as differences between the two surfaces larger than those expected as a result of random error within each model at a given confidence level. In this study, where planimetric registration between the DEMs is robust simple, vertical differencing is appropriate. In cases where there is significant misregistration a surface-normal differencing approach should be considered. The threshold (T) for defining gross error was calculated from (Taylor, 1997):

$$T = t_{ax}\sqrt{\sigma_p^2 + \sigma_c^2} \quad (1)$$

where: t_{ax} is the two-tailed critical t-test value at a given confidence level (x); σ_p and σ_c are the theoretical precisions of the generated and existing DEMs respectively. For the generated DEMs, the theoretical precision is related to the ground pixel resolution, the dimensions of each pixel in the

TABLE 1. ACQUISITION AND PROCESSING DETAILS FOR THE TWO AERIAL PHOTOGRAPH DATASETS USED IN THIS STUDY

Dataset	Archival	Commissioned
Source	INFOTERRA 2004	ULM 2005
Camera Type	RC20	Zeiss Jenna LMK
Focal Length *(f)*	153 mm	152 mm
Flying Height *(H)*	2600 m	2100 m
Base Length	1247 m	920 m
Aerial Photo-scale	1:14000	1:11000
Scanning Resolution (d_e)	14 μm	8 μm
Scanner Type	Zeiss Scai	Vexcel, Ultrascan 5000
Mean Signal-to-noise Ratio	123.9 & 124.7	123.5 & 123.6
Ground Pixel Resolution (d_o)	0.198 m	0.114 m
Interior Orientation RMSE	7.22 & 6.67 μm	6.26 & 6.38 μm
Exterior Orientation RMSE	0.9821 pixels	0.9983 pixels

TABLE 2. DETAILS OF THE THREE CHECK DATASETS TESTED IN THIS STUDY

	IFSAR (Interferometric Synthetic Aperture Radar)	OS (Ordnance Survey Profile)	SRTM (Shuttle Radar Topography Mission)
Source	Intermap	Ordnance Survey	NASA
Coverage	United Kingdom	United Kingdom	Global
Availability	Free to UK NERC funded academics £36/m² commercially	Free to UK academics £0.17/m² commercially	Free
Collection Method	Airborne IFSAR	Analytical Photogrammetry	Satellite IFSAR
Resolution	5 m	10 m	50 m
Stated Precision	1.5 m	2.5 m	16 m
Reference	Intermap (2004)	Ordnance Survey (2005)	Rabus *et al.* (2003)

object space (d_o) and the image scale, such that it can be estimated *a priori* using:

$$p \approx d_o = \frac{d_e}{f/H} \tag{2}$$

where: d_e is the size of the pixel in the image space, equivalent to the scanning resolution; H is the distance of the sensor from the ground surface and f is the focal length of the sensor. If stereo-matching using area based correlation is adopted, the expected precision is often degraded by a factor of the correlation window size. In this case the default correlation window is 7 × 7 pixels so the theoretical precision (σ_p) is $7p$. The stated precisions for each existing DEM (Table 2) were used as *a priori* estimates of their precision. Table 3 shows the estimated precision for each dataset and the matrix of the calculated thresholds that define gross error for each set of air photos with each check dataset. Thresholds were set to define points as gross error with 95 percent ($x = 0.05$) and 99 percent ($x = 0.01$) confidence. Elevations with differences greater than the 95 percent and 99 percent confidence thresholds were labeled as gross error and used to generate error maps (e.g., Figures 1 and 2). The total area affected by gross error at each confidence interval and the standard deviation of the error (STD) between datasets was calculated.

Parameter Optimization

We used three sets of parameters to assess the effect of different levels of user interference in the stereo-matching process on surface precision. These were: (a) the default set contained within the chosen software system, (b) the "best pre-defined" set recommended by the manufacturer of the software system for the type of terrain being examined, and (c) a set that we optimized. Initial testing of the stereo-matching algorithm identified four parameters (Table 4) that had a significant effect on the resultant digital elevation data. These parameters were varied using a constrained sensitivity analysis. Each parameter was perturbed in isolation across a representative range (Table 4). The effect of changing each parameter on the generated surface was quantified using the DEM comparison method. Parameters were then set to their optima and re-perturbed to identify any change in their sensitivity or optimum value resulting from parameter interaction.

TABLE 3. THE ESTIMATED PRECISION, IN METERS, OF CHECK DATASETS (Σ_c); THEORETICAL PRECISION OF PHOTOGRAMMETRIC DATASETS (Σ_p) AND THRESHOLDS FOR GROSS ERROR (T) CALCULATED USING EQUATION 1 FOR EACH COMBINATION OF DATASETS AT EACH CONFIDENCE INTERVAL

		Precision (σ)	SRTM 16.00	OS 2.50	IFSAR 1.50
Archival	95%	1.67	31.53	5.89	4.39
	99%		46.33	8.65	6.45
Commissioned	95%	0.77	31.40	5.13	3.31
	99%		46.13	7.54	4.86

Independent Check Data for Assessing the Effects of Optimization

In order to assess the impact of using existing DEM data, we acquired independent check data using a real-time kinematic differential global positioning system (DGPS), precise to better than ± 0.05 m. We collected 1,000 GPS points along a series of transects across the study site. These were used to determine the mean error (ME) and the standard deviation of error (STD) for each generated dataset.

Results

Figures 1 and 2 show the error maps for the default, best pre-defined and optimum parameter sets. The maps have been generated using each of the available existing datasets and show where the elevation differences between the generated and existing DEM are larger than those expected due to random error in the two surfaces (gross errors). The location and areal extent of the error patches change with: (a) the parameter set used, (b) the existing DEM data used for comparison, and (c) the type of imagery. The IFSAR and Ordnance Survey (OS) data identify similar areas of error in each surface. In both cases, optimization using these existing data can be used to reduce the number of locations and areal extent of gross error to a much greater degree than is possible by adopting a pre-defined terrain sensitive parameter set. Use of SRTM data resulted in the same errors being identified in some locations. However, a large area classed as gross error in the west of the study site appears to be the result of poor topographic representation in the SRTM model. Also, the larger expected random error (σ_c) within the SRTM data means that some areas classified as gross error with 99 percent confidence using the IFSAR and OS datasets fall

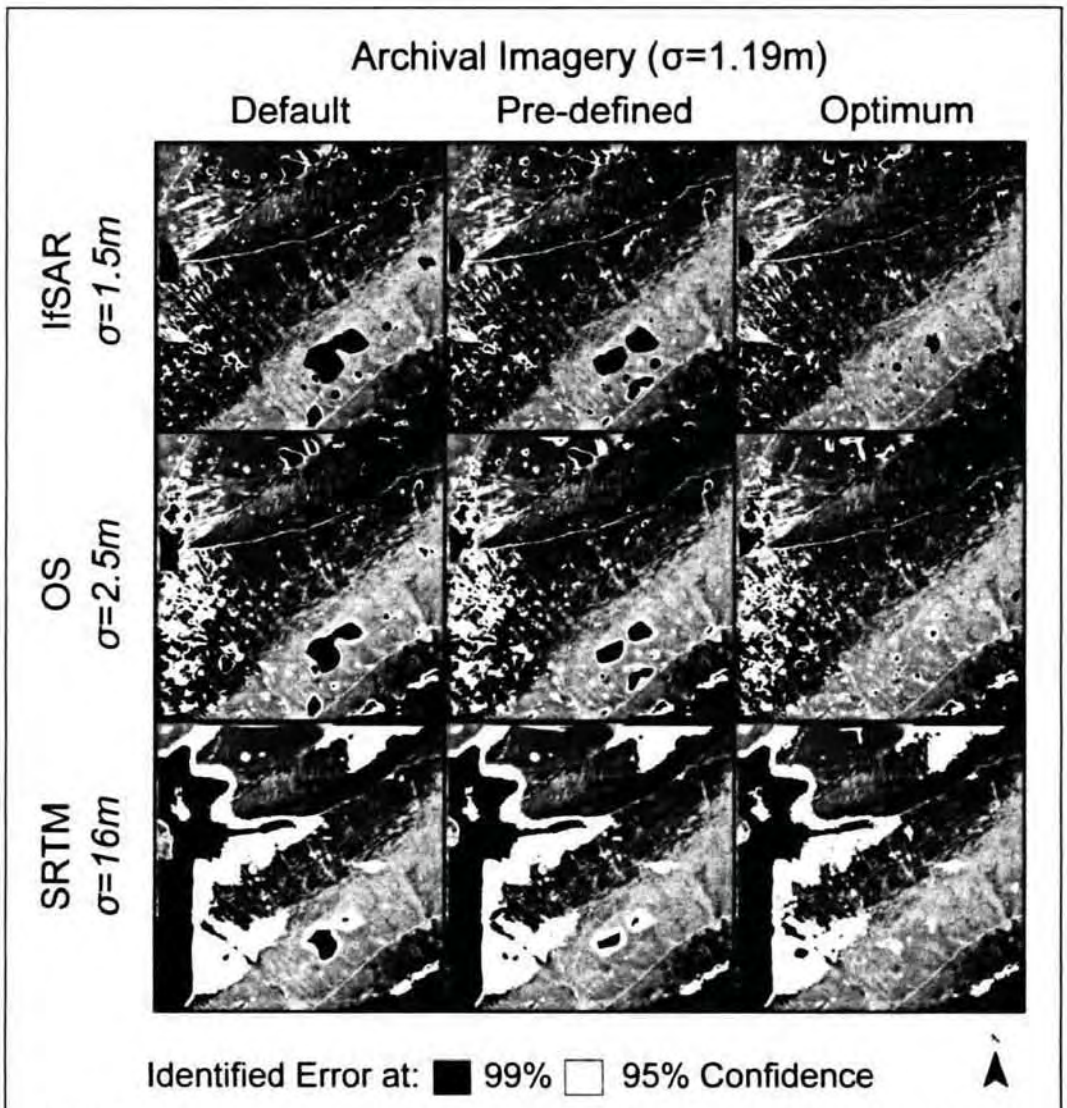

Figure 1. A matrix of DEM comparison error maps (1 km^2) for elevation models generated from archival imagery. The check datasets are shown along the vertical axis and the parameter set used to generate the surfaces is indicated on the horizontal axis, σ is the theoretical precision for each dataset.

within the range of expected random error for the SRTM data. Comparison of Figures 1 and 2 shows that these conclusions are independent of the source of data used for DEM generation. Extant Optimization had most impact on the commissioned imagery, reducing the area of gross error to 0.3 percent, although its effects were still notable in the archival imagery where the area covered by error was reduced to less than 2 percent (Table 5).

Figures 3 and 4 compare the errors identified in Figures 1 and 2 with those measured independently using DGPS. For both the IFSAR and OS data, there is a strong correlation between the error identified by comparing a generated DEM with existing data and the error identified by comparing a generated DEM with measured check data. In other words, existing DEM data can be a useful substitute for specially-collected check data. However, this is not the case for SRTM data where the correlation between identified errors and check data is weak ($R^2 < 0.12$ and slope <0.07 in all cases). Extant Optimization successfully reduced the gross error in models generated from both sets of imagery. Figure 5 shows the cumulative frequency distributions of error for each generated DEM as compared with the check data. DEMs from both the archival and commissioned imagery have high magnitude negative residuals for surfaces generated with default parameters. These result in large standard deviations of error (Table 5). For the commissioned imagery, the negative residuals are reduced by adopting the best pre-defined parameter set. However, this has little effect on error in the archival imagery DEM. The negative residuals in both surfaces are further reduced after Extant Optimization yielding improvements in the standard deviation of error. Howev as very few GPS points fall in areas identified as gross errors on the DEM comparison maps, these results probably represent a conservative estimate of both the magnitude of gross error in the default parameter surfac and the extent to which it is reduced by parameter optimization. This is a common problem with check dat which is rarely available at the resolution necessary to rigorously interrogate a DEM (Lane *et al.*, 2000) and emphasizes the value of using existing DEM data in anal ses of this kind.

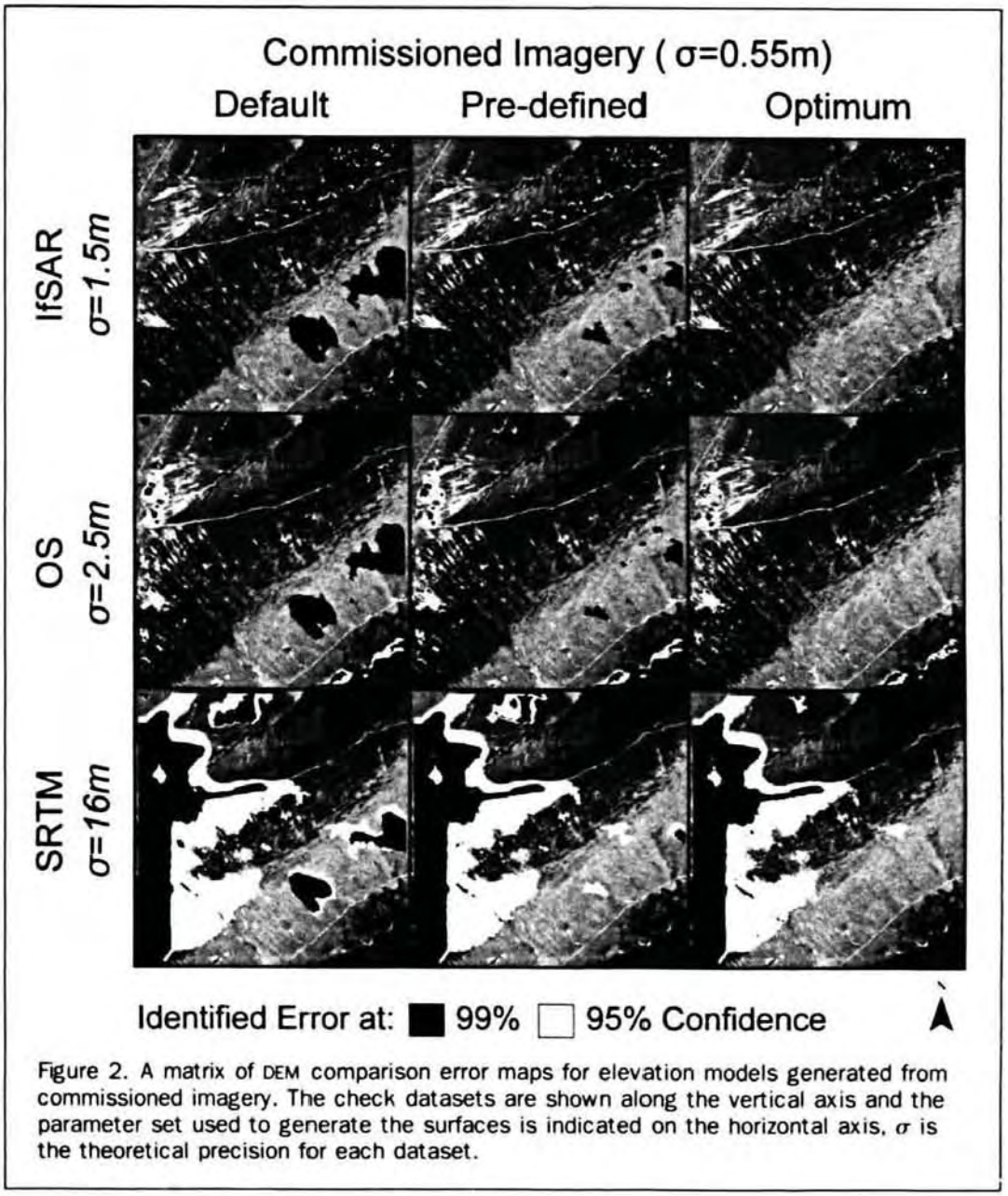

Figure 2. A matrix of DEM comparison error maps for elevation models generated from commissioned imagery. The check datasets are shown along the vertical axis and the parameter set used to generate the surfaces is indicated on the horizontal axis, σ is the theoretical precision for each dataset.

Figure 6 illustrates the influence of the quality of existing data on the parameter optimization process for the commissioned and archival imagery. Parameter variations are expressed as a percentage of their reasonable range (Table 4; Leica Geosystems, 2003). Resultant changes in the area affected by gross error are expressed as a percentage of the gross error in the default DEM. Points exceeding the 99 percent confidence threshold were considered gross errors. The 99 percent threshold appeared to respond to parameter change in a more stable manner, and we could be more confident that identified points were a result of error in the generated DEM rather than poor surface representation in the existing DEM. The commissioned and archival DEMs behaved differently in response to parameter variation and returned different optimum parameter sets. However, there were some similarities for both the commissioned and

TABLE 4. THE FOUR USER-DEFINED PARAMETERS WITHIN THE IMAGE-MATCHING ALGORITHM THAT ARE VARIED IN THIS STUDY, THEIR ABBREVIATION, DEFAULT VALUE AND REASONABLE RANGE (LANE ET AL., 2000; LEICA GEOSYSTEMS, 2003)

Parameter	Description	Default	Range
s_x	Search window size along epipolar line	21	15–70+
s_y	Search window size across epipolar line	3	1–5
Cw	Size of the window used to correlate pixels	7	5–15
CC	Correlation coefficient limit for accepting a match	0.6	0.5–0.7

TABLE 5. ELEVATION ERROR STATISTICS CALCULATED FROM GPS CHECK DATA: MEAN ERROR (ME), STANDARD DEVIATION OF ERROR (STD), IN METERS, AND THE PERCENTAGE OF EACH DEM TAGGED BY THE COMPARISON METHOD AS GROSS ERROR USING EACH EXISTING DATASET

		% Error Points				
		IFSAR	OS	SRTM	ME	STD
Archival	Default	5.6	3.4	6.1	1.10	1.71
	Best predefined	3.3	1.9	6.0	1.15	1.59
	Optimum	2.0	0.6	5.4	1.37	1.22
Commissioned	Default	4.9	4.0	7.8	0.8	3.15
	Best predefined	2.2	1.6	5.9	1.50	1.19
	Optimum	0.3	0.3	5.8	1.70	0.94

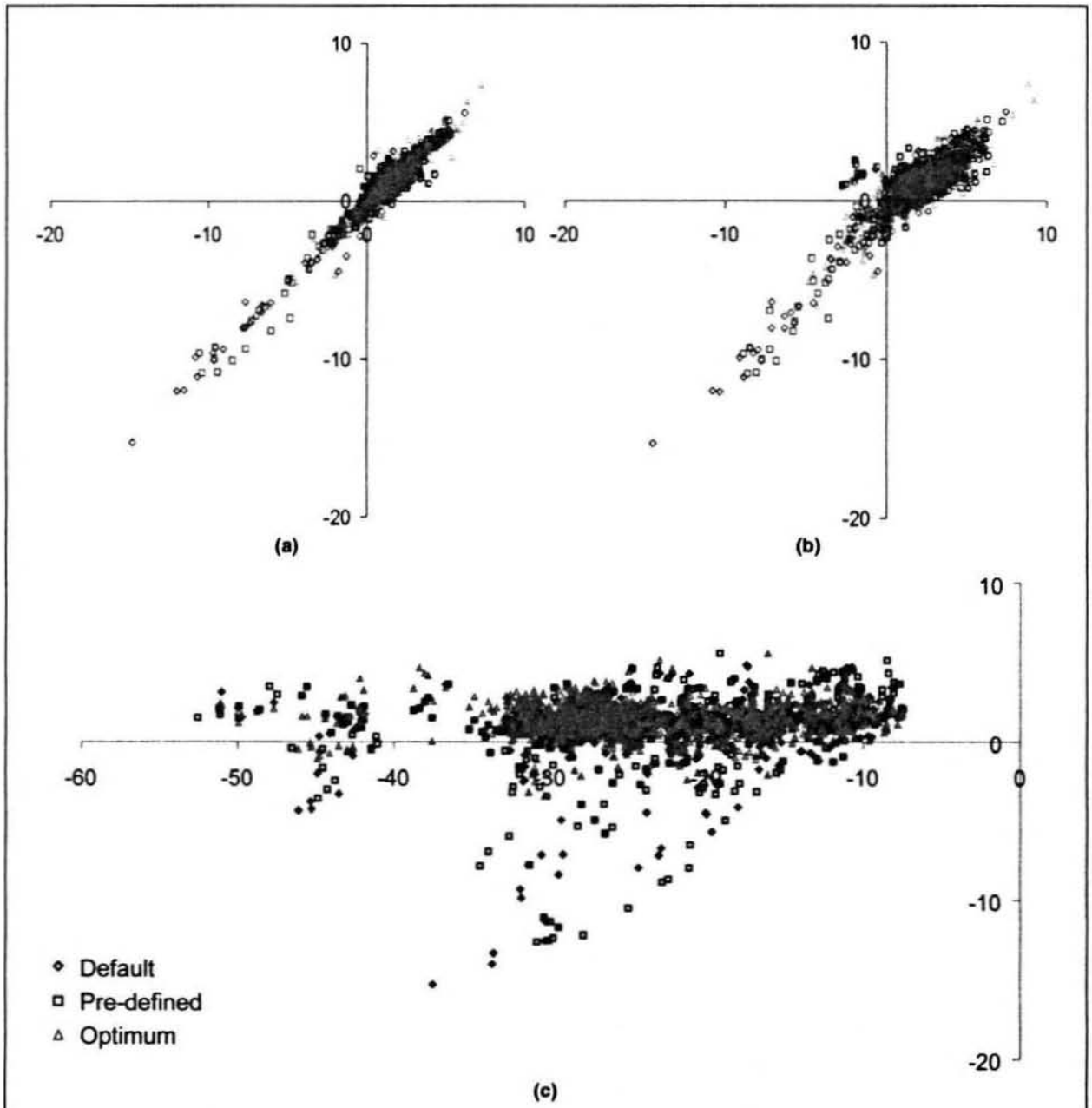

Figure 3. Plots of predicted errors plotted on the *x*-axis (using (a) IFSAR, (b) Ordnance Survey, and (c) SRTM data) against observed elevation errors plotted on the *y*-axis (using GPS check points) for elevation models generated from the archival imagery. In each case plots show the relationship for surfaces generated using default, best pre-defined and optimum parameters.

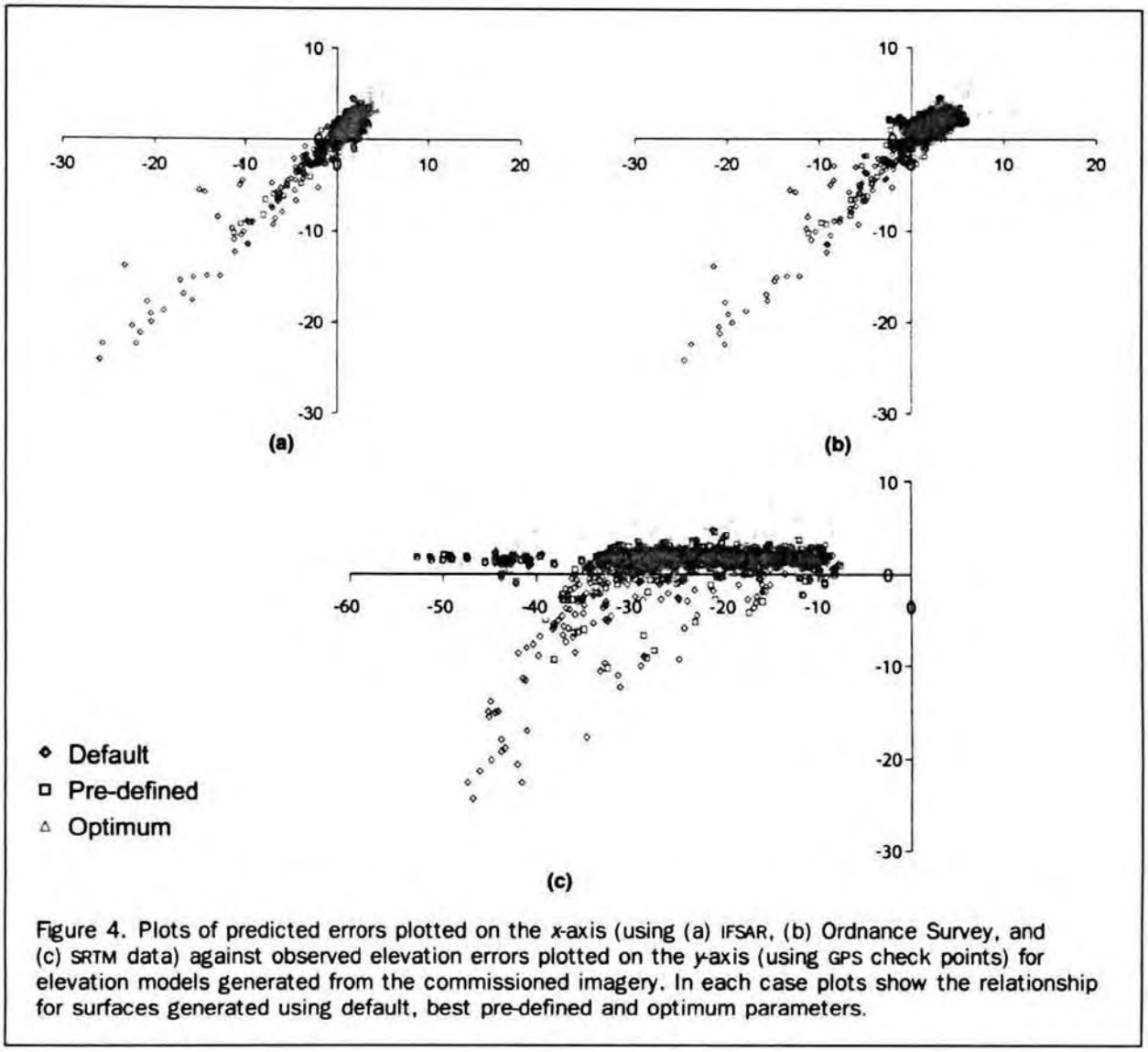

Figure 4. Plots of predicted errors plotted on the *x*-axis (using (a) IFSAR, (b) Ordnance Survey, and (c) SRTM data) against observed elevation errors plotted on the *y*-axis (using GPS check points) for elevation models generated from the commissioned imagery. In each case plots show the relationship for surfaces generated using default, best pre-defined and optimum parameters.

archival data. Gross error was most sensitive to the size of the search window along the epipolar line (sw_x) and the correlation coefficient and least sensitive to the size of the search window across the epipolar line (sw_y) and the correlation window. In all cases error declined exponentially as the size of the search window (x) was increased, although the form of the decline varied in each case.

For the commissioned data (Figure 6), the default correlation window and correlation coefficient were also their optima. The parameter sensitivity plots using IFSAR and OS data were very similar and identified the same optimum parameters. The SRTM data displays similar trends to the other check datasets. However, the magnitude of the variation in error is considerably reduced as a result of the large area mis-classified as gross error due to poor topographic representation in the SRTM DEM. The parameterization process for the archival data was more complex. The default parameter values were rarely optimal and there was more variability between optimum values defined by each check dataset. Under initial sensitivity analysis the archival imagery appeared more sensitive to the correlation window size and less sensitive to the correlation coefficient than the commissioned imagery. When the parameters were perturbed from their optima, parameter interaction effects not encountered during parameterization of the commissioned data were identified. Optimum values were difficult to obtain from the SRTM data, but clearly identifiable using the IFSAR and OS data, which produced almost identical relationships between parameter variation and gross error. Table 4 gives the default and pre-defined parameter values and Table 6 the optimum values for each dataset. The variability in optimum parameter values, even when the surface of interest remains constant, highlights the importance of parameter optimization using existing data.

Two alternative methods of identifying photogrammetric error, the Failure Warning Model and that of Westaway *et al.* (2003), were assessed using two different parameter sets as examples (Table 7). Both these techniques use two DEMs generated from the same image pair, identifying error as areas where the surfaces differ significantly. Figure 7 shows suspect areas identified by the Failure Warning Model (FWM) for two parameter sets (Gooch and Chandler, 2001). The DEM comparison error maps identify large patches of error in the surface generated using parameter set 1 and very little error in the surface using parameter set 2. The Failure Warning Model successfully identifies many of the areas of error resulting from parameter set 1. However, because the maps are created by differencing the generated surfaces from the default, which itself has considerable gross error, it is impossible to disentangle the source of the errors. It is possible to identify areas in the DEMs that are sensitive to the matching parameters but not to identify whether these

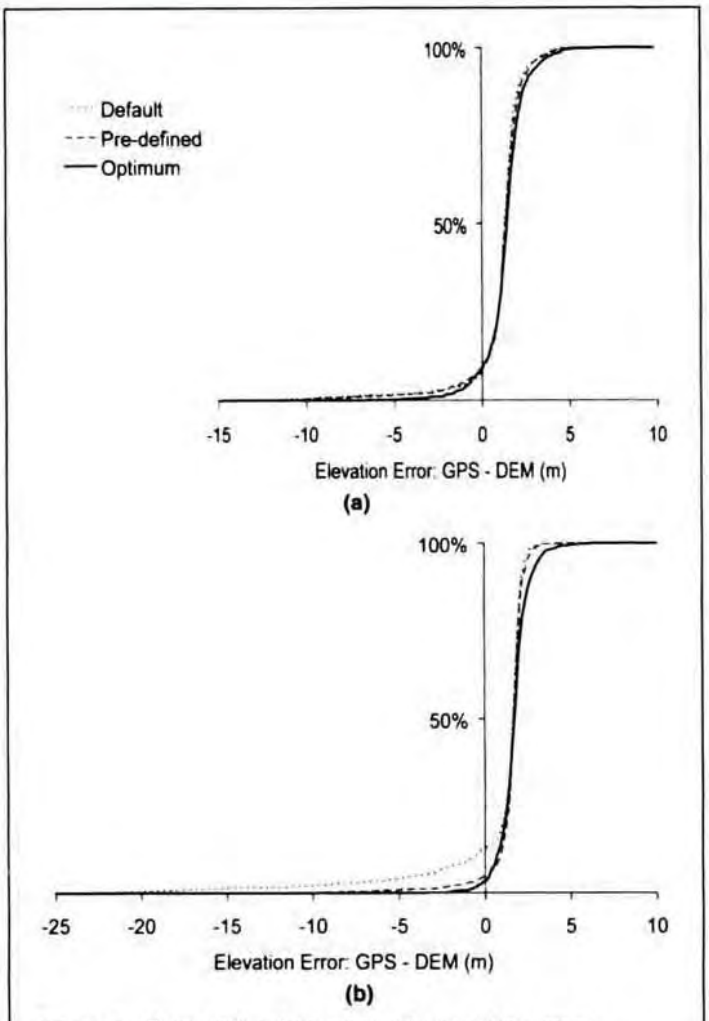

Figure 5. Cumulative frequency of elevation errors, calculated from GPS check data, for surfaces generated from (a) archival, and (b) commissioned imagery using default, best pre-defined and optimum parameter-sets.

areas are in error. Thus, this does not lead to the improvement of the acquired surface, but simply a dataset of points that are suspect and which, if removed, will need to be replaced by interpolated points. Further, in some areas affected by gross error in both surfaces, no significant difference in elevation is observed with parameter variation. These areas are not considered suspect by the FWM but contain significant gross error. The method described by Westaway *et al.* (2003) identifies large areas of error within parameter set 1 and fewer errors in parameter set 2. However, when the component surfaces used to derive the error maps are analyzed using the DEM comparison method it is clear that significant gross errors exist in both the 5 m and 1 m DEMs. For this reason the Westaway *et al.* (2003) method does not identify error within a generated surface nor does it identify changes in DEM error in response to changing parameter values.

Discussion

Our first hypothesis was that Extant Optimization could produce a higher quality DEM by identifying better parameter sets than those already available. Results from this study have shown that in some cases the terrain sensitive pre-defined parameter set recommended within the software can reduce the area affected by failures in the matching algorithm. However, the parameter optimization results show that the interaction between matching parameters, their importance and their optimum values vary on a case-by-case basis. In such an environment, pre-defined parameter sets are unable to capture the optimum values for these parameters. Extant Optimization captures these optima, identified as clear minima in the area of mapped gross error as parameter values are varied across their reasonable range. DEM generation with these identified optimum matching parameters effectively reduces the frequency of high magnitude errors within the surface improving its precision (Figure 5). In order to be confident in Extant Optimization we must be confident that the error maps created by differencing generated surfaces from an alternative dataset effectively identify gross error in the photogrammetric DEMs. The strong correlation between errors predicted using IFSAR or OS data and those observed (Figures 3 and 4) suggests that this is the case.

Some caution is required in the application of this method. First, in some cases, optimization appears to increase the frequency of low magnitude error within the optimized surface (Figure 5). However, these errors can be easily removed by the application of conventional filters (e.g., Felicisimo, 1994; Walker and Willgoose, 2006). Therefore, it represents less of a problem for surface representation than the patches of gross error, which would cause error to propagate through the DEM if filters were applied to smooth the surface. Second, the reliance of Extant Optimization on a comparison between independent datasets requires that there is no significant change due to geomorphological processes (i.e., erosion and deposition) between the two collection dates. There will be a relationship here between the magnitude of error that the technique is capable of identifying and the magnitude of change in the study area. In areas where change might be expected *a priori*, the user defined threshold for gross error could be adjusted to account for this.

Our second hypothesis, that Extant Optimization can be used to compensate for situations where image content is less than ideal, generated mixed results. It is true that optimizing the matching parameters can improve the precision of the archival DEM to a level comparable with or better than that available from the commissioned imagery using the best available parameter set within the software. In fact, DEMs generated with poor matching parameters appear insensitive to image content. Using default parameters, the precision of the DEM generated from commissioned photography was considerably worse than that generated from archival imagery. However, as the matching parameters improve, the DEM errors become more sensitive to image content and the difference in precision between the two datasets is maintained through the parameter optimization.

Hypothesis three asserts that the effectiveness of Extant Optimization depends on the type of existing DEM data available. The use of three different check datasets in parallel has shown that errors within the check data can reduce the reliability with which errors can be predicted. This affects the ability of the DEM comparison method to identify gross errors within the photogrammetric DEM and a result has an impact on parameter optimization. However the decline in effectiveness with error in the existing data does not appear to be linear. IFSAR and OS data are both effective check datasets for parameterization. The degradation in reliability for OS data with respect to IFSAR data is small (Figures 3 and 4), with little difference between identified optimum parameter sets for each (Figure 6). This suggests that pre-existing cartographic maps represent a suitable dataset for mapping gross errors and optimizing matching parameters. This is an important finding, since

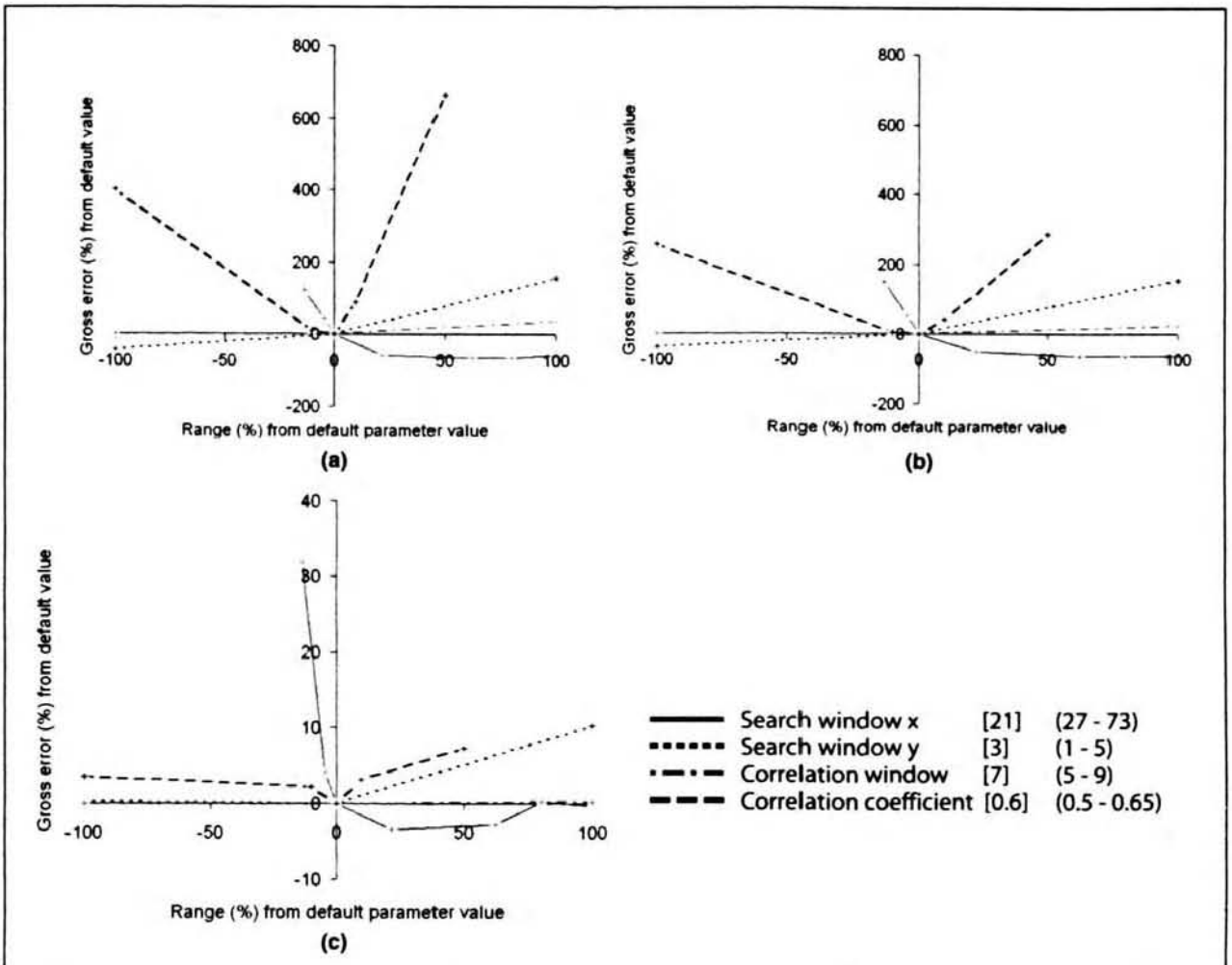

Figure 6. Sensitivity of commissioned DEM error to matching parameters. Parameters were perturbed from their default value (in square brackets) over their reasonable range (in brackets). The resulting change in gross error is measured as a percentage of the gross error with default values. Gross error is estimated with error maps using data from: (a) IFSAR, (b) Ordnance Survey, and (c) SRTM.

these datasets are extensively available at relatively low cost making this technique widely applicable. The reduced reliability of SRTM data (Figures 3 and 4) suggests that this data is less suitable for use in DEM comparison error mapping. It performed reasonably well at parameter optimization, identifying the same trends as the other datasets. However, the identifiablility of optima was reduced as a result of poor surface representation in the SRTM data and the identified optimum parameters often differed from those identified using IFSAR or OS data. The problems experienced using SRTM data for Extant Optimization highlight a key relationship between the resolution of the existing data, the magnitude of the error threshold and the topographic roughness of the surface. At their original resolution (50 m for SRTM) existing data represent a simplified form of the surface. This data must then be interpolated to match the resolution of the photogrammetric data, using bilinear interpolation the new value is determined based on a weighted distance average of the four nearest input cell centers. If the topographic variability over an area equiva-

TABLE 6. OPTIMUM MATCHING PARAMETER VALUES FOR THE ARCHIVAL AND COMMISSIONED IMAGERY, ESTIMATED USING EXTANT OPTIMIZATION WITH THE THREE EXISTING ELEVATION DATASETS IFSAR, OS, AND SRTM

		Search Window (x)	Search Window (y)	Correlation Window	Correlation Coefficient
Archival	IFSAR	39	3	7	0.55
	OS	39	3	7	0.55
	SRTM	49	3	7	0.7
Commissioned	IFSAR	54	1	7	0.6
	OS	54	1	7	0.6
	SRTM	39	3	7	0.6

TABLE 7. MATCHING PARAMETER VALUES WITH ERROR STATISTICS, MEAN ERROR (ME) AND STANDARD DEVIATION OF ERROR (STD), CALCULATED BY COMPARISON WITH GPS CHECKPOINTS, FOR THE SURFACES USED TO COMPARE ERROR MAPS

Parameter set	Search Window (x)	Search Window (y)	Correlation Window	Correlation Coefficient	ME (m)	STD (m)
1	21	3	5	6	1.39	1.60
2	39	3	7	0.55	1.37	1.22

lent to the resolution is greater than the defined threshold for gross error then the ability of this dataset to identify errors in the photogrammetric DEM will be considerably reduced.

Finally, our analysis of hypothesis four suggests that by using existing datasets, areas identified in error maps can be confidently considered error. This is not the case using self-referenced check data (Gooch and Chandler, 2001; Westaway *et al.*, 2003). An analysis of these methods illustrates their key limitation: error present in both surfaces means that the surface responsible for the mapped error cannot be confidently identified. As a result, suspect areas can be located but reductions in error as a result of changes in the matching parameters cannot be seen. This study confirms our fourth hypothesis, that although these techniques remain capable of identifying areas of uncertainty they are not suitable as indicators of the amount of error in a surface and in particular for use as error indicators within a parameterization framework.

Stereo-matching performance is constantly improving, both as a result of improvements in image quality and in the algorithms used to match them. In particular, the new generation of digital metric cameras have a superior radiometric performance to imagery scanned from analogue sensors and early digital cameras. This should allow more successful stereo-matching and lead to fewer gross errors. The fusion of enhanced stereo-matching with other rapidly developing technologies, such as lidar, represents a promising future for topographic remote sensing. Here, however, we use lower technology datasets ubiquitous, both worldwide and through the recent past, to demonstrate a truly generic method of optimizing image-matching.

Conclusions

Current methods of identifying error or uncertainty within photogrammetric DEMs rely on comparison between two surfaces generated from the same imagery. These methods can identify suspect areas but cannot indicate the provenance of the error within them. As a result, they are unable to confidently identify the location or extent of gross errors in such elevation data. DEM comparison effectively predicts the location and extent of gross errors within DEMs derived using stereo-matching. With such data it is possible to improve the quality of the surface generated by stereo-matching through optimizing the matching parameters. This process can be effectively

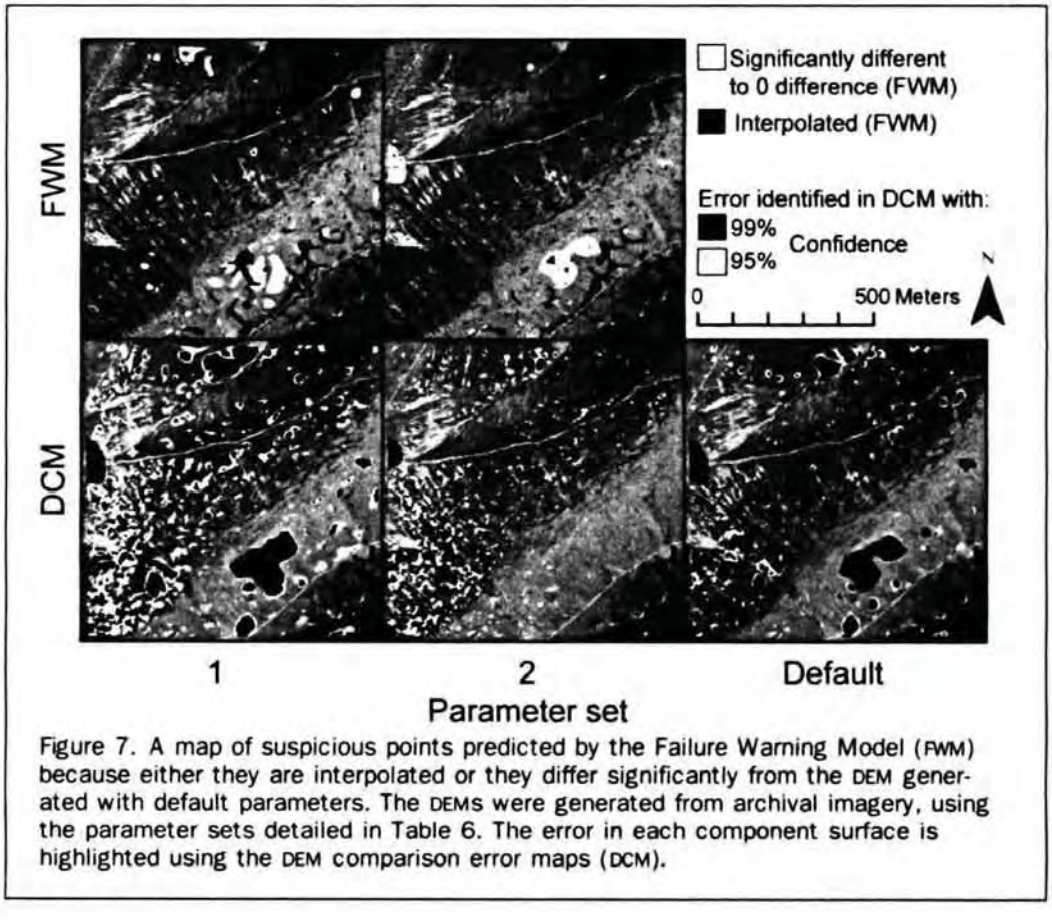

Figure 7. A map of suspicious points predicted by the Failure Warning Model (FWM) because either they are interpolated or they differ significantly from the DEM generated with default parameters. The DEMs were generated from archival imagery, using the parameter sets detailed in Table 6. The error in each component surface is highlighted using the DEM comparison error maps (DCM).

performed using elevation data derived from traditional cartographic maps. SRTM data is of limited use, because although it identifies similar trends, different optima are determined during parameter optimization. We conclude that Extant Optimization is a potentially valuable methodology for improving elevation data at the point of generation. It can be applied easily using existing data that is widely and cheaply available and has the potential to improve the precision of elevation models based upon stereo-matching by over 30 percent relative to other alternatives (e.g., terrain sensitive parameter sets).

These conclusions are drawn from a case study using DEMs generated from true vertical aerial photos in a rural setting of moderate relief. Further application of the method suggested here should be evaluated in contrasting environments to establish the generality of these findings.

Acknowledgments

This research was supported by NERC Small Grant NE/D/521481/1 awarded to Jeff Warburton. David G. Milledge was funded by NERC Ph.D. Studentship NER/S/A/2004/12248. OS data were supplied by Ordnance Survey/EDINA under Crown Copyright/database right 2007. NEXTMAP data were supplied by Intermap Technologies through the NERC Earth Observation Data Centre.

References

Butler, J.B., S.N. Lane, and J.H. Chandler, 1998. Assessment of DEM quality for characterizing surface roughness using close range digital photogrammetry, *The Photogrammetric Record*, 16(92):271–291.

Felicisimo, A.M., 1994. Parametric statistical method for error-detection in digital elevation models, *ISPRS Journal of Photogrammetry and Remote Sensing*, 49(4):29–33.

Gooch, M.J., and J.H. Chandler, 2001. Failure prediction in automatically generated digital elevation models, *Computers & Geosciences*, 27(8):913–920.

Gooch, M.J., J.H. Chandler, and M. Stojic, 1999. Accuracy assessment of digital elevation models generated using the ERDAS Imagine Orthomax digital photogrammetric system, *The Photogrammetric Record*, 16(93):519–531.

Intermap, 2004. *Intermap Product Handbook and Quick Start Guide*, Intermap Technologies, Englewood, Colorado, 118 p.

Intermap, 2007. Nextmap 3D mapping program, URL: *http://intermap.com/right.php/pid/1/sid/318*, Intermap Technologies, Englewood, Colorado, (last date accessed: 03 December 2008).

Lane, S.N., T.D. James, and M.D. Crowell, 2000. Application of digital photogrammetry to complex topography for geomorphological research, *The Photogrammetric Record*, 16(95):793–821.

Lane, S.N., S.C. Reid, R.M. Westaway, and D.M. Hicks, 2004. Remotely sensed topographic data for river channel research: The identification, explanation and management of error, *Spatial Modelling of the Terrestrial Environment* (R.E.J. Kelly, N.A. Drake and S.L. Barr, editors), Wiley, pp. 157–174.

Leica Geosystems, 2003. *Leica Photogrammetry Suite, Orthobase and Orthobase Pro User's Guide*, Leica Geosystems, Atlanta, Georgia, 490 p.

Lim, M., D.N. Petley, N.J. Rosser, R.J. Allison, A.J. Long, and D. Pybus, 2005. Combined digital photogrammetry and time-of-flight laser scanning for monitoring cliff evolution, *The Photogrammetric Record*, 20(110):109–129.

Ordnance Survey, 2005. Land-Form PROFILE Plus technical sheet, URL:*http://www.ordnancesurvey.co.uk/oswebsite/products/landformprofileplus/pdf/LandFormPROFILE%20Plus.pdf*, Ordnance Survey, Southampton, UK, (last date accessed: 03 December 2008).

Rabus, B., M. Eineder, A. Roth, and R. Bamler, 2003. The shuttle radar topography mission - A new class of digital elevation models acquired by spaceborne radar, *ISPRS Journal of Photogrammetry and Remote Sensing*, 57(4):241–262.

Taylor, J., 1997. *Introduction to Error Analysis, The Study of Uncertainties in Physical Measurements*, University Science Books, New York, 327 p.

Walker, J.P., and G.R. Willgoose, 2006. A comparative study of Australian cartometric and photogrammetric digital elevation model accuracy, *Photogrammetric Engineering & Remote Sensing*, 72(7):771–779.

Westaway, R.M., S.N. Lane, and D.M. Hicks, 2003. Remote survey of large-scale braided, gravel-bed rivers using digital photogrammetry and image analysis, *International Journal of Remote Sensing*, 24(4):795–815.

Yanites, B.J., R.H. Webb, P.G. Griffiths, and C.S. Magirl, 2006. Debris flow deposition and reworking by the Colorado River in Grand Canyon, Arizona, *Water Resources Research*, 42(11):W11411.

(Received 30 August 2007; accepted 03 December 2007; revised 21 December 2007)